Searching for a prodrome for rheumatoid arthritis in the primary care record

Muller, Sara; Hider, Samantha; Machin, Annabelle; Stack, Rebecca; Hayward, Richard A.; Raza, Karim; Mallen, Christian

DOI:
10.1016/j.semarthrit.2018.06.008
10.1093/rheumatology/key075.539

Searching for a prodrome for rheumatoid arthritis in the primary care record: A case-control study in the Clinical Practice Research Datalink

Sara Muller[1], Samantha Hider[1,2], Annabelle Machin[1], Rebecca Stack[3], Richard A Hayward[1], Karim Raza[4,5], Christian Mallen[1]

Corresponding author: Sara Muller. David Weatherall Building, Keele University, Keele, Staffordshire, ST5 5BG, UK. T: +44 (0)1782 734842

[1]Research Institute for Primary Care & Health Sciences, Keele University, Keele, UK

[2]Haywood Academic Rheumatology Centre, Haywood Hospital, Stoke-on-Trent, UK

[3]College of Business Law & Social Sciences, Nottingham Trent University, Nottingham, UK

[4]Institute of Inflammation and Ageing, Arthritis Research UK Rheumatoid Arthritis Pathogenesis Centre of Excellence and MRC Arthritis Research UK Centre for Musculoskeletal Ageing Research, University of Birmingham, Birmingham, UK

[5] Department of Rheumatology, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK

ABSTRACT

Background

Rheumatoid arthritis (RA) has articular and non-articular manifestations. Early, intensive treatment has substantial benefit for both. This requires patients be identified as soon as symptoms develop.

Objectives

To determine whether selected signs and symptoms can be identified in the primary care records of patients prior to a formal diagnosis of RA being made and, if so, how early they can be identified.

Methods

A case-control study was constructed within the UK Clinical Practice Research Datalink (CPRD). 3577 individuals with 'definite' RA, were matched to 14287 individuals without inflammatory arthritis. An index date was established (i.e. date general practitioner (GP) first appeared to suspect RA). Rates of consultation and consultations for suspected early RA symptoms were compared in cases and controls in the two years prior to the index date using conditional logistic regression, adjusted for number of consultations.

Results

The mean (standard deviation) age of participants was 58.8 (14.5) years and 66.8% were female. Rates of any consultation were significantly higher in RA cases than in controls for at least two years prior to the index date. Cases were more likely to have a pre-diagnosis coded consultation for joint, and particularly hand symptoms (aOR 11.44 (9.60, 13.63)), morning stiffness (8.10 (3.54, 18.5)), carpal tunnel syndrome (4.57 (3.54, 5.88)) and other non-articular features.

Conclusions

In patients who develop RA, GP consultation rates are higher for at least two years prior to the first recorded suspicion of RA. This study highlights symptoms that should raise a GP's index of suspicion for RA.

KEY WORDS: Rheumatoid arthritis; Diagnosis; Primary care

INTRODUCTION

Rheumatoid arthritis (RA) causes joint pain, stiffness and damage and can lead to excess morbidity and mortality. It has a prevalence in the UK of around 0.67% [1]. It is known that early, intensive treatment can increase the likelihood of remission and reduce long term joint damage and comorbidities [2,3].

Delay in making a diagnosis of RA, and therefore in treating it, can occur at a number of points in the patient journey [3-6]; first in the patient recognising their symptoms and seeking help from primary care, second in the primary care physician recognising the potential for a diagnosis of RA and making a referral to a rheumatology specialist, and third in seeing a rheumatologist and starting appropriate treatment. Work has been ongoing to understand the causes of patient delay [7-12], which has comprised in-depth studies of the symptoms reported by patients prior to their diagnosis of RA [13-16]. These symptoms have included problems with joints, fatigue, weakness [13], muscle cramps, psychological distress, and loss of motor control [14].

Primary care delay continues to be a significant contributor to overall diagnostic delay for people with RA [17]. This could be because GPs are not aware of the need to refer quickly [18], or because they find it difficult to identify 'red flag' symptoms, for example because of co-existing musculoskeletal conditions [19,20]. We hypothesised that we would be able to identify signs and symptoms in the coded part of the medical record that increase the likelihood of a future RA diagnosis, increasing GPs awareness and facilitating a more rapid referral.

We present a case-control study to assess the association of clinical features reported by patients in the earliest phases of RA with future diagnosis of the condition using the UK Clinical Practice Research Datalink (CPRD).

MATERIALS AND METHODS

Data source: the Clinical Practice Research Datalink

The CPRD is an anonymised source of routinely collected primary care health records covering approximately 6.9% of the UK's population. It is broadly representative of this population in terms of age, sex and ethnicity [21]. The data exist in coded form and include details of symptoms, diagnoses and prescriptions. Clinical data are assigned Read codes (the hierarchical clinical coding system currently used in UK primary care) by the GP. Data are collected for clinical purposes and so it can be assumed that anything the GP considers relevant might be coded, regardless of whether any clinical action was required as a result. The coding of data has been shown to be accurate for a range of conditions [22]. The CPRD assigns an 'up-to-standard' date for when a practice has a high enough quality of coding to be used for research. We use only up-to-standard data in this study.

Definition of rheumatoid arthritis

Previous work in the General Practice Research Database (predecessor to CPRD) developed an algorithm to identify individuals with RA [23]. The algorithm combines Read codes for RA with prescription information and potential alternative diagnoses to create a case definition that is specific, but not overly sensitive. It has been updated to allow for the inclusion of new Read codes and the use of biological disease modifying anti-rheumatic drugs (DMARD) [24].

Analysis sample

All individuals with a first Read code for RA in the CPRD between 2007 and 2012 were identified and the algorithm to define RA was applied [24]. Those who met this specific definition of RA were then matched to four individuals of the same sex and from the same practice who were born in the same 3-year time interval who did not have a Read code for any inflammatory arthropathy up until the time of the case's first RA code, in order to form a case-control study.

The index date: first indication of RA in the records

The date at which the first RA Read code was recorded may not accurately reflect the date at which the GP first suspected RA in the patient [25,26], as he/she may wait until diagnosis is confirmed by a specialist before entering this diagnostic code. As a proxy for the date of first clinical suspicion of RA, an index date was defined. Based on previous work in the CPRD [27,28], this index date was taken to be the earliest of the first RA code, or other code from the Read code subchapter N04 (rheumatoid arthritis and other inflammatory arthropathy), the date of the first prescription of a DMARD or first referral to rheumatology in the three years preceding the first RA Read code (Figure 1).

Early signs and symptoms

Signs and symptoms that may precede a diagnosis of RA were identified from the literature [13-16] and consultation with experts. Whilst some of the signs/symptoms described in the literature were not possible to define within the medical record (e.g. muscle burning as one would feel after exercise), other signs/symptoms were clearly defined syndromes and conditions that could be specified and studied in more detail (e.g. carpal tunnel syndrome, shoulder pain).

The final set of the signs/symptoms included is given in Tables 2 (articular) and 3 (non-articular). Lists of Read codes to define each sign/symptom, the concepts they represent and the process to achieve the lists are available at [29].

Where a sign/symptom was not recorded, it was assumed that the individual did not experience it, rather than data being missing.

Statistical analyses

Rates of consultation

A consultation was defined as a day on which a Read coded contact with the practice was made. Where there were multiple contacts/Read codes on the same day, only one was included in the consultation rate analysis. Monthly rates of consultations in the 2-year period before the index date

were estimated and compared in cases and controls using incidence rate ratios with 95% confidence intervals (CI).

Signs and symptoms associated with RA

We investigated signs/symptoms in the 2-year period prior to the index date. Cumulative time periods were defined at 1, 6, 12 and 24 months prior to the index date. Each period included previous periods (Figure 1). Additionally, we considered the period 12 to 24 months before the index date to allow comparison with the period 0 to 12 months before the index date. In these analyses all Read coded contacts were considered, even when multiple codes were entered on the same date.

Conditional logistic regression was used to assess the association between signs/symptoms and case control (RA) status, allowing for the matched design in each of the time periods described above. Results are presented as unadjusted odds ratios and then adjusted for the rate of consultations (as defined above) in the period in question. All values are presented as odds ratios (OR) with 95% CI.

Data management and analyses were conducted in Stata 14.2 [30].

Approval for this study was granted by the Independent Scientific Advisory Committee of the CPRD (reference 13_126).

RESULTS

Sample

Between 2007 and 2012, 4161 people were identified with a first Read code for RA. Of these, 3577 met the criteria for RA using the algorithm described above [24] and were matched to 14287 controls (Table 1). The mean (standard deviation) age of the both cases and controls was 58.8 (14.5)

144 years and 66.8% were female (Table 1). Current and previous smoking were more common in cases
145 than controls.
146
147 **Numbers of consultations**
148 For two years prior to the index date, the overall consultation rate was significantly higher in each
149 month in cases than in controls (Figure 2) (incidence rate ratio 1.22 (1.21, 1.22)). This increase
150 became more pronounced in the 6 months before the index date and in the final month before,
151 cases consulted at 2.68 (95% CI 2.61, 2.76) times the rate of controls (mean: 2.39 consultations per
152 person).
153
154 **Signs and symptoms preceding a diagnosis of RA**
155 Articular symptoms
156 All articular signs/symptoms were associated with a future diagnosis of RA in all time periods in both
157 unadjusted and adjusted analyses, with the exception of jaw pain which was not significantly
158 associated with RA following adjustment for the number of consultations in the 0 to 1-month period
159 (Table 2). In the 0 to 1-month period, joint symptoms (adjusted OR (95% CI), 14.82 (12.48, 17.60)),
160 and specifically hand problems (61.07 (31.58, 118.10)), were strongly associated with the
161 development of RA. Palindromic rheumatism occurred only in cases in the 12 months preceding the
162 index date. The strength of all associations was lessened by adjustment for the number of
163 consultations (except for jaw pain in the 0 to 6-month period) and associations were generally
164 stronger for consultations closer to the index date.
165
166 Non-articular symptoms
167 In the 0 to 1-month period there were four non-articular signs/symptoms that had large (odds
168 ratio≥6) unadjusted associations with development of RA (morning stiffness, muscle pain: 13.83
169 (5.11, 37.42) and carpal tunnel syndrome: 2.96 (1.38, 6.34)). All remained strongly and significantly

associated after adjustment (Table 3). Morning stiffness was recorded in 14 cases (0.39%), but not in any controls, hence an estimate of the strength of association could not be made. Unintentional weight loss was not significantly associated with RA in the month before the index date, but was in all other time periods (except adjusted analysis in 12 to 24-month period (Supplementary tables)). In all time periods, there was a significant unadjusted association between fatigue and development of RA, but this association was attenuated and not significant after adjustment for number of consultations. No association was seen with sleep problems or flu-like illness (Table 3). Psychological problems were significantly associated with a higher odds of RA in all unadjusted analyses (except 0 to 24 months before the index date, where association was positive, but not significant), but significantly associated with a decreased odds of RA after adjustment.

Comparison of consultation 0-12 and 12-24 months before the index date

Comparison of the associations of signs/symptoms during the periods 0 to 12-months and 12 to 24-months before the index date suggested that signs/symptoms grouped together (Supplementary tables). A similar pattern of association was seen in both time periods for fatigue, altered sensations, postnatal occurrence of RA, weakness, psychological problems and carpal tunnel syndrome. There was no unadjusted association between sleep and RA in either time period and only after adjustment in the 12 to 24-month period for falling. Flu-like symptoms were associated with RA in the 12 to 24-month period (adjusted analyses only (1.46 (1.04, 2.07)), but not closer to the index date. All articular signs/symptoms and the remaining non-articular signs/symptoms were more strongly associated with RA in the 0 to 12-month period than in the 12 to 24-month period. This was particularly noticeable for hand symptoms (aOR 0 to 12-months: 23.75 (18.49, 30.51); 12 to 24-months: 2.70 (2.05, 3.56)), morning stiffness (0 to 12-months: 9.72 (3.84, 24.60); 12 to 24-months: 0.93 (0.08, 10.64)) and muscle pain (0 to 12-months: 3.15 (2.22, 4.47); 12 to 24-months: 1.22 (0.75, 1.97)).

DISCUSSION

The rate of consultations increases rapidly in the period before the index date in those with RA compared to controls, and key signs and symptoms are recorded at a higher rate before an RA diagnosis. In the final month before the index date, these include all joint symptoms, but particularly those involving the hand, and the non-articular symptoms morning stiffness, muscle pain and carpal tunnel syndrome. In longer periods before the index date, there is also an increase in the recording of other features such as unintentional weight loss. Other symptoms reported by patients in previous studies (e.g. fatigue, cramping, poor sleep) [13-16], showed less clear associations.

The strengths of our study include the large sample size and use of a validated definition of RA [24]. Whilst it could be argued that the definition had good specificity at the expense of sensitivity, the exclusion of controls with any record of inflammatory arthropathy should reassure that there was no contamination of the control group with potential cases. The data for this study were taken from a high quality database containing a representative sample of individuals in UK primary care. As such, the results should be generalizable to other primary care settings. Despite its strengths, this study also has some weaknesses. First, multiple statistical testing, which could result in false positive associations. Second, we do not know the thought processes of the GPs who coded the consultations and how this might have affected our findings. This question cannot be answered in routinely collected data, but would require in-depth interviews with GPs as to their views and clinical practice. This in itself may prove difficult, as an individual GP will see a new case of RA only rarely and may not be able to report what action they would take [20].

We adjusted our findings for the total number of consultations (days with ≥1 Read coded consultation) in order to adjust for ascertainment/surveillance bias, whereby the presence of the patient in the surgery makes it easier for the GP to identify and code signs/symptoms. This adjustment for number of consultations attenuated the association of a number of signs/symptoms

(physical functioning, cramps, weakness and restless legs) with RA. This may suggest that these signs/symptoms are more common in those with RA, but are only recorded in those who attend more frequently. A similar process may explain the change in direction of association with psychological problems when adjusting for number of consultations: people with coded psychological problems consult more frequently and it may be that controls receive psychological codes, but RA cases receive codes for physical symptoms because these take priority for the GP. Due to small numbers, non-significant associations in the final month before the index date should be interpreted with caution, as they may well represent a type II statistical error, especially where the signs/symptom was associated with RA in longer time periods and the absolute estimate of the size of association is similar across time periods (e.g. unintentional weight loss). However, it could be that these symptoms are simply more common at an earlier stage in the pre-clinical picture of RA and become less commonly reported or over-shadowed by other symptoms in the final weeks before the GP suspects RA.

Previous literature has described the symptoms patients report before a diagnosis of RA [13-16], and there is a feeling among rheumatologists that they know what symptoms they expect to see in early RA. However, to our knowledge this is the first paper to consider whether these signs and symptoms occur in the primary care record, whether they are more common in those who later received a diagnosis of RA than in those who do not and how long before RA is suspected by the GP these signs and symptoms are present.

Within this study, classical features of RA such as hand pain and stiffness were more frequently coded within the primary care record, and were seen more frequently up to 2 years before the index date. However, musculoskeletal symptoms in regions not traditionally associated with early RA (e.g. neck and shoulder pain) were also reported more frequently by patients who eventually developed RA. Joint symptoms, particularly in the hands, and other well-recognised non-articular features

should raise the index of suspicion of RA in patient presenting in primary care, particularly when accompanied by a general increase in patient contact with the primary care. However, GPs should also be aware that these features have low specificity and only a small proportion of patients with these symptoms will go on to receive a diagnosis of RA. For example, whilst we have confirmed that people with RA are more likely to have hand symptoms, an RA outcome is seen in only a minority of patients that have hand symptoms recorded in primary care. Further studies will be needed to investigate what other symptoms/signs increase the likelihood of an RA diagnosis.

Other early symptoms reported by patients such as falls and sleep problems did not show any association with RA. This may represent a true lack of association, or it may be that either patients did not report these symptoms to the GP or that GPs did not code them, especially if they did not fit with the GP's concept of what is important.

The association of flu-like symptoms with RA only in the 12 to 24-month period may suggest that rather than being part of an RA prodrome, flu-like symptoms may be a marker of an insult on the immune system that reflects the phase of immune tolerance breakage [31].

The next steps should be to identify groups of symptoms that constitute a prodrome of RA and at the same time educate GPs as to the key symptoms that may indicate RA prior to the cardinal symptoms of morning stiffness and hand symptoms that they already appear to recognise, although further work is needed to refine the specificity of these common symptoms.

In the future, it may be possible to create automated electronic alerts for the GP within the records system that highlight the risk for an individual patient when certain codes are entered. This already happens for example to alert the GP to the possibility of sepsis.

What our study was not able to do was to identify new signs or symptoms from the record that may occur at a higher rate in those who go on to receive an RA diagnosis than in those who do not; to do

so would have required an alternative methodological approach to identify patterns in consultation that were not defined by code lists (e.g. 32]).

CONCLUSION

We have provided definitive evidence of the presence of some key features of early RA in the primary care medical record prior to the GP appearing to recognise the condition. Primary care professionals should be aware of the range of articular and non-articular features, specifically hand symptoms, muscle pain, carpal tunnel syndrome and unintentional weight loss, accompanied by an increased rate of consultation, as potentially forming a prodromal syndrome for RA. Increased awareness of these symptoms combined with education on the need for early referral could facilitate earlier treatment of RA, increasing the likelihood of remission and reducing long term joint damage and comorbidities.

CONFLICT OF INTERESTS

KR reports personal fees from BMS, personal fees from Abbvie, grants from Pfizer, personal fees from Pfizer, personal fees from UCB, outside the submitted work. The other authors report no conflicts of interest.

ACKNOWLEDGEMENTS

The authors are grateful to Dr Alyshah Abdul Sultan and Mrs Rebecca Whittle for their help in processing the CPRD data.

Funding

SM and AM are funded by the National Institute of Health Research School for Primary Care Research. CDM is funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care West Midlands, the NIHR School for Primary Care Research and a NIHR Research Professorship in General Practice (NIHR-RP-2014-04-026). KR is

supported by the NIHR Birmingham Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Abhishek A, Doherty M, Kuo CF, Mallen CD, Zhang W, Grainge MJ. Rheumatoid arthritis is getting less frequent-results of a nationwide population-based cohort study. *Rheumatology (Oxford)* 2017; **56:** 736-44.
2. Moura CS, Abrahamowicz M, Beauchamp ME, Lacaille D, Wang Y, Boire G, et al. Early medication use in new-onset rheumatoid arthritis may delay joint replacement: results of a large population-based study. *Arthritis Res Ther* 2015; **17:** 197.
3. van der Linden MP, le Cessie S, Raza K, van der Woude D, Knevel R, Huizinga TW, et al. Long-term impact of delay in assessment of patients with early arthritis. *Arthritis Rheum* 2010; **62:** 3537-46.
4. Barhamain AS, Magliah RF, Shaheen MH, Munassar SF, Falemban AM, Alshareef MM, et al. The journey of rheumatoid arthritis patients: a review of reported lag times from the onset of symptoms. *Open Access Rheumatol* 2017; **9:** 139-50.
5. Kumar K, Daley E, Carruthers DM, Situnayake D, Gordon C, Grindulis K, et al. Delay in presentation to primary care physicians is the main reason why patients with rheumatoid arthritis are seen late by rheumatologists. *Rheumatology (Oxford)* 2007; **46:** 1438–40.
6. Feldman DE, Bernatsky S, Houde M, Beauchamp ME, Abrahamowicz M. Early consultation with a rheumatologist for RA: does it reduce subsequent orthopaedic use? *Rheumatology (Oxford)* 2013; **52:** 452–9.
7. Simons G, Lumley S, Falahee M, Kumar K, Mallen CD, Stack RJ, et al. The pathway to consultation for rheumatoid arthritis: exploring anticipated actions between the onset of symptoms and face-to-face encounter with a healthcare professional. *BMC Musculoskelet Disord* 2017; **18:** 258.

8. Simons G, Mason A, Falahee M, Kumar K, Mallen CD, Raza K, et al. Qualitative Exploration of Illness Perceptions of Rheumatoid Arthritis in the General Public. *Musculoskeletal Care* 2017; **15:** 13-22.
9. Mølbaek K, Hørslev-Petersen K, Primdahl J. Diagnostic Delay in Rheumatoid Arthritis: A Qualitative Study of Symptom Interpretation Before the First Visit to the Doctor. *Musculoskeletal Care* 2016; **14:** 26-36.
10. Van der Elst K, De Cock D, Vecoven E, Arat S, Meyfroidt S, Joly J, et al. Are illness perception and coping style associated with the delay between symptom onset and the first general practitioner consultation in early rheumatoid arthritis management? An exploratory study within the CareRA trial. *Scand J Rheumatol* 2016; **45:** 171-8.
11. Molina E, Del Rincon I, Restrepo JF, Battafarano DF, Escalante A. Association of socioeconomic status with treatment delays, disease activity, joint damage, and disability in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2015; **67:** 940-6.
12. Simons G, Mallen CD, Kumar K, Stack RJ, Raza K. A qualitative investigation of the barriers to help-seeking among members of the public presented with symptoms of new-onset rheumatoid arthritis. *J Rheumatol* 2015; **42:** 585-92.
13. Stack RJ, Sahni M, Mallen CD, Raza K. Symptom complexes at the earliest phases of rheumatoid arthritis: a synthesis of the qualitative literature. *Arthritis Care Res (Hoboken)* 2013; **65:** 1916-26.
14. Stack RJ, van Tuyl LH, Sloots M, van de Stadt LA, Hoogland W, Maat B, et al. Symptom complexes in patients with seropositive arthralgia and in patients newly diagnosed with rheumatoid arthritis: a qualitative exploration of symptom development. *Rheumatology (Oxford)* 2014; **53:** 1646-53.
15. van Tuyl LH, Stack RJ, Sloots M, van de Stadt LA, Hoogland W, Maat B, et al. Impact of Symptoms on Daily Life in People at Risk of Rheumatoid Arthritis. *Musculoskeletal Care* 2016; **14:** 169-73.

16. Newsum EC, de Waal MW, van Steenbergen HW, Gussekloo J, van der Helm-van Mil AH. How do general practitioners identify inflammatory arthritis? A cohort analysis of Dutch general practitioner electronic medical records. *Rheumatology (Oxford)* 2016; **55:** 848-53.
17. Healthcare Quality Improvement Partnership. A patient and public guide to the National Clinical Audit for Rheumatoid and Early Inflammatory Arthritis 1st Annual Report 2015 (Data collection: 1 February 2014 – 30 April 2015). Available at: https://www.hqip.org.uk/public/cms/253/625/19/456/WEB-BSR%20Patient%20Report%202015.pdf?realName=SMFSaA.pdf&v=0 Accessed December 2017
18. Patient view on behalf of the National Audit Office. People with Rheumatoid Arthritis, their carers, and the NHS. A national survey of patients with RA and their carers. 2009. Available at: https://www.nao.org.uk/wp-content/uploads/2009/07/0809823_Patient_survey.pdf Accessed December 2017
19. Lard LR, Huizinga TW, Hazes JM, Vliet Vlieland TP. Delayed referral of female patients with rheumatoid arthritis. *J Rheumatol* 2001; **28:** 2190-2.
20. Meyfroidt S, Stevens J, De Lepeleire J, Westhovens R, De Cock D, Van der Elst K, et al. A general practice perspective on early rheumatoid arthritis management: A qualitative study from Flanders. *Eur J Gen Pract* 2015; **21:** 231-7.
21. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44:** 827-36.
22. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69:** 4-14.
23. Thomas SL, Edwards CJ, Smeeth L, Cooper C, Hall AJ. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? *Arthritis Rheum* 2008; **59:** 1314-21.

24. Muller S, Hider SL, Raza K, Stack RJ, Hayward RA, Mallen CD. An algorithm to identify rheumatoid arthritis in primary care: a Clinical Practice Research Datalink study. *BMJ Open* 2015; **5:** e009309.
25. Ford E, Carroll J, Smith H, et al. What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text. *BMJ Open* 2016; **6:** e010393.
26. Ford E, Nicholson A, Koeling R, Davies K, Koeling R, Petersen I, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol* 2013; **13:** 105.
27. Nicholson A, Ford E, Davies KA, Smith HE, Rait G, Tate AR, et al. Optimising use of electronic health records to describe the presentation of rheumatoid arthritis in primary care: a strategy for developing code lists. *PLoS One* 2013; **8:** e54878.
28. Tate AR, Martin AG, Murray-Thomas T, Anderson SR, Cassell JA. Determining the date of diagnosis--is it a simple matter? The impact of different approaches to dating diagnosis on estimates of delayed care for ovarian cancer in UK primary care. *BMC Med Res Methodol* 2009; **9:** 42.
29. Medical Record Data Research. 2018. keele.ac.uk/mrr. Accessed 23 Mar 2018.
30. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP, 2015.
31. Tracy A, Buckley CD, Raza K. Pre-symptomatic autoimmunity in rheumatoid arthritis: when does the disease start? *Semin Immunopathol* 2017; **39:** 423-35.
32. Zhou SM, Fernandez-Gutierrez F, Kennedy J, Cooksey R, Atkinson M, Denaxas S, et al. Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis. *PLoS One* 2016; **11:** e0154515.

Corresponding author: Sara Muller, Research Institute for Primary Care & Health Sciences, David Weatherall Building, Keele University, Keele, UK. s.muller@keele.ac.uk

Second corresponding author: Samantha Hider, Research Institute for Primary Care & Health Sciences, David Weatherall Building, Keele University, Keele, UK. s.hider@keele.ac.uk

406 **Figure Legends**

407 Figure 1 Schematic representation of data set

408 Figure 2 Consultation rates (per person year) in cases and controls prior to index date